

PCT

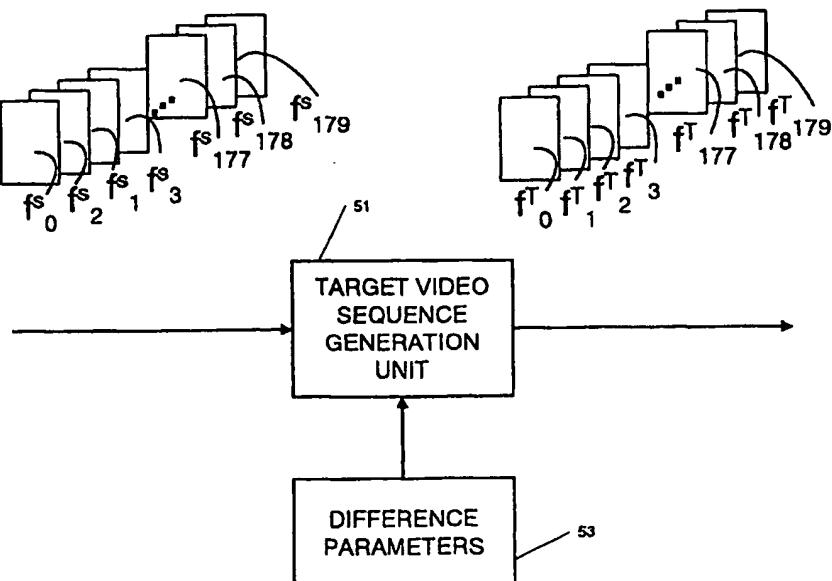
WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : <b>G06T 15/70</b>	A1	(11) International Publication Number: <b>WO 00/17820</b> (43) International Publication Date: 30 March 2000 (30.03.00)
(21) International Application Number: <b>PCT/GB99/03161</b>		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 22 September 1999 (22.09.99)		
(30) Priority Data: 9820633.7 22 September 1998 (22.09.98) GB		
(71) Applicant ( <i>for all designated States except US</i> ): ANTHROPICS TECHNOLOGY LIMITED [GB/GB]; c/o Createc, Ealing Studios, Ealing Green, London W5 5EP (GB).		
(72) Inventors; and		
(75) Inventors/Applicants ( <i>for US only</i> ): BEREND, Andrew, Louis, Charles [GB/GB]; Createc, Ealing Studios, Ealing Green, London W5 5EP (GB). WILLIAMS, Mark, Jonathan [GB/GB]; Createc, Ealing Studios, Ealing Green, London W5 5EP (GB).		
(74) Agents: BERESFORD, Keith, Denis, Lewis et al.; Beresford & Co., 2-5 Warwick Court, High Holborn, London WC1R 5DJ (GB).		

(54) Title: GRAPHICS AND IMAGE PROCESSING SYSTEM



(57) Abstract

An image and graphics processing system is provided which can automatically generate an animated sequence of images of a deformable object by combining a source video sequence with a target image. The system may be used to simulate hand-drawn and computer-generated animations of characters.

BEST AVAILABLE COPY

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

GRAPHICS AND IMAGE PROCESSING SYSTEM

The present invention relates to a method of and apparatus for graphics and image processing. The invention has particular, although not exclusive, relevance to the image processing of a sequence of source images to generate a sequence of target images. The invention has applications in computer animation and in moving pictures.

Realistic facial synthesis is a key area of research in computer graphics. The applications of facial animation include computer games, video conferencing and character animation for films and advertising. However, realistic facial animation is difficult to achieve because the human face is an extremely complex geometric form.

The paper entitled "Synthesising realistic facial expressions from photographs" by Pighin et al published in Computer Graphics Proceedings Annual Conference Series, 1998, describes one technique which is being investigated for generating synthetic characters. The technique extracts parts of facial expressions from input images and combines these with the original image to generate different facial expressions. The system then uses a morphing technique to animate a change between different facial expressions. The generation of an animated sequence therefore involves the steps of identifying a required sequence of facial expressions (synthetically generating any if necessary) and then morphing between each expression to generate the animated sequence. This technique is therefore relatively complex and requires significant operator input to control the synthetic generation of new facial expressions.

One embodiment of the present invention aims to provide an alternative technique for generating an animated video sequence. The technique can be used to generate realistic facial animations or to generate simulations of hand drawn facial animations.

According to one aspect, the present invention provides an image processing apparatus comprising: means for receiving a source sequence of frames showing a first object; means for receiving a target image showing a second object; means for comparing the first object with the second object to generate a difference signal; and means for modifying each frame of the sequence of frames using said difference signal to generate a target sequence of frames showing the second object.

This aspect of the invention can be used to generate 2D animations of objects. It may be used, for example, to animate a hand-drawn character using a video clip of, for example, a person acting out a scene. The technique can also be used to generate animations of other objects, such as other parts of the body and animals.

A second aspect of the present invention provides a graphics processing apparatus comprising: means for receiving a source sequence of three-dimensional models of a first object; means for receiving a target model of a second object; means for comparing a model of the first object with the model of the second object to generate a difference signal; and means for modifying each model in the sequence of models for the first object using said difference signal to generate a target sequence of models of the second object.

According to this aspect, three-dimensional models of,

for example, a human head can be modelled and animated in a similar manner to the way in which the two-dimensional images were animated.

5 The present invention also provides methods corresponding to the apparatus described above.

Exemplary embodiments of the present invention will now be described with reference to the accompanying drawings  
10 in which:

Figure 1 is a schematic block diagram illustrating a general arrangement of a computer system which can be programmed to implement the present invention;

15 Figure 2a is a schematic illustration of a sequence of image frames which together form a source video sequence;

20 Figure 2b is a schematic illustration of a target image frame which is to be used to modify the sequence of image frames shown in Figure 2a;

25 Figure 3 is a block diagram of an appearance model generation unit which receives some of the image frames of the source video sequence illustrated in Figure 2a together with the target image frame illustrated in Figure 2b to generate an appearance model;

30 Figure 4 is a flow chart illustrating the processing steps employed by the appearance model generation unit shown in Figure 3 to generate the appearance model;

35 Figure 5 is a flow diagram illustrating the steps involved in generating a shape model for the training images;

Figure 6 shows a head having a number of landmark points placed over it;

5       Figure 7 illustrates the processing steps involved in generating a grey level model from the training images;

10      Figure 8 is a flow chart illustrating the processing steps required to generate the appearance model using the shape and grey level models;

15      Figure 9 shows the head shown in Figure 6 with a mesh of triangles placed over the head;

15      Figure 10 is a plot showing a number of landmark points surrounding a point;

20      Figure 11 is a block diagram of a target video sequence generation unit which generates a target video sequence from a source video sequence using a set of stored difference parameters;

25      Figure 12 is a flow chart illustrating the processing steps involved in generating the difference parameters;

25      Figure 13 is a flow diagram illustrating the processing steps which the target video sequence generation unit shown in Figure 11 performs to generate the target video sequence.

30      Figure 14a shows three frames of an example source video sequence which is applied to the target video sequence generation unit shown in Figure 11;

35      Figure 14b shows an example target image used to generate a set of difference parameters used by the target video

sequence generation unit shown in Figure 11;

5       Figure 14c shows a corresponding three frames from a target video sequence generated by the target video sequence generation unit shown in Figure 11 from the three frames of the source video sequence shown in Figure 14a using the difference parameters generated using the target image shown in Figure 14b;

10      Figure 14d shows a second example of a target image used to generate a set of difference parameters for use by the target video sequence generation unit shown in Figure 11; and

15      Figure 14e shows the corresponding three frames from the target video sequence generated by the target video sequence generation unit shown in Figure 11 when the three frames of the source video sequence shown in Figure 14a are input to the target video sequence generation 20 unit together with the difference parameters calculated using the target image shown in Figure 14d.

25      Figure 1 is a block diagram showing the general arrangement of an image processing apparatus according to an embodiment of the present invention. The apparatus comprises a computer 1 having a central processing unit (CPU) 3 connected to a memory 5 which is operable to store a program defining the sequence of operations of the CPU 3 and to store object and image data used in 30 calculation by the CPU 3.

35      Coupled to an input port of the CPU 3 there is an input device 7, which in this embodiment comprises a keyboard and a computer mouse. Instead of, or in addition to the computer mouse, another position sensitive input device

(pointing device) such as a digitiser with associated stylus may be used.

A frame buffer 9 is also provided and is coupled to the  
5 CPU 3 and comprises a memory unit (not shown) arranged to store image data relating to at least one image, for example by providing one (or several) memory location(s) per pixel of the image. The value stored in the frame buffer for each pixel defines the colour or intensity of  
10 that pixel in the image. In this embodiment, the images are represented by 2-D arrays of pixels, and are conveniently described in terms of cartesian coordinates, so that the position of a given pixel can be described by a pair of x-y coordinates. This representation is  
15 convenient since the image is displayed on a raster scan display 11. Therefore, the x-coordinate maps to the distance along the line of the display and the y-coordinate maps to the number of the line. The frame buffer 9 has sufficient memory capacity to store at least  
20 one image. For example, for an image having a resolution of 1000 x 1000 pixels, the frame buffer 9 includes  $10^6$  pixel locations, each addressable directly or indirectly in terms of pixel coordinates x,y.  
25 In this embodiment, a video tape recorder (VTR) 13 is also coupled to the frame buffer 9, for recording the image or sequence of images displayed on the display 11. A mass storage device 15, such as a hard disc drive, having a high data storage capacity is also provided and  
30 coupled to the memory 5. Also coupled to the memory 5 is a floppy disc drive 17 which is operable to accept removable data storage media, such as a floppy disc 19 and to transfer data stored thereon to the memory 5. The memory 5 is also coupled to a printer 21 so that  
35 generated images can be output in paper form, an image

input device 23 such as a scanner or video camera and a modem 25 so that input images and output images can be received from and transmitted to remote computer terminals via a data network, such as the internet.

5

The CPU 3, memory 5, frame buffer 9, display unit 11 and mass storage device 13 may be commercially available as a complete system, for example as an IBM compatible personal computer (PC) or a workstation such as the Spark station available from Sun Microsystems.

10

A number of embodiments of the invention can be supplied commercially in the form of programs stored on a floppy disc 19 or other medium, or as signals transmitted over a data link, such as the internet, so that the receiving hardware becomes reconfigured into an apparatus embodying the present invention.

15

In this embodiment, the computer 1 is programmed to receive a source video sequence input by the image input device 23 and to generate a target video sequence from the source video sequence using a target image. In this embodiment, the source video sequence is a video clip of an actor acting out a scene, the target image is an image of a second actor and the resulting target video sequence is a video sequence showing the second actor acting out the scene. The way in which this is achieved in this embodiment will now be described with reference to Figures 2 to 11.

20

Figure 2a schematically illustrates the sequence of image frames ( $f^s$ ) making up the source video sequence. In this embodiment, there are 180 source image frames  $f^s_0$  to  $f^s_{179}$  making up the source video sequence. In this embodiment, the frames are black and white images having 500 x 500

35

pixels, whose value indicates the luminance of the image at that point. Figure 2b schematically illustrates the target image  $f^T_1$ , which is used to modify the source video sequence. In this embodiment, the target image is also a black and white image having 500 x 500 pixels, describing the luminance over the image.

In this embodiment, an appearance model is generated for modelling the variations in the shape and grey level (luminance) appearance of the two actors' heads. In this embodiment, the appearance of the head and shoulders of the two actors is modelled. However, for simplicity, in the remaining description reference will only be made to the heads of the two actors. This appearance model is then used to generate a set of difference parameters which describe the main differences between the heads of the two actors. These difference parameters are then used to modify the source video sequence so that the actor in the video sequence looks like the second actor. The modelling technique employed in the present embodiment is similar to the modelling technique described in the paper "Active Shape Models - Their Training and Application" by T.F. Cootes et al, Computer Vision and Image Understanding, Vol. 61, No. 1, January pp 38-59, 1995, the contents of which are incorporated herein by reference.

#### TRAINING

In this embodiment, the appearance model is generated from a set of training images comprising a selection of frames from the source video sequence and the target image frame. In order for the model to be able to regenerate any head in the video sequence, the training images must include those frames which have the greatest variation in facial expression and 3D pose. In this

embodiment, seven frames ( $f_{s_3}^s$ ,  $f_{s_{26}}^s$ ,  $f_{s_{34}}^s$ ,  $f_{s_{47}}^s$ ,  $f_{s_{98}}^s$ ,  $f_{s_{121}}^s$  and  $f_{s_{162}}^s$ ) are selected from the source video sequence as being representative of the various different facial expressions and poses of the first actor's face in the 5 video sequence. As shown in Figure 3, these training images are input to an appearance model generation unit 31 which processes the training images in accordance with user input from the user interface 33, to generate the appearance model 35. In this embodiment, the user 10 interface 33 comprises the display 11 and the input device 7 shown in Figure 1. The way in which the appearance model generation unit 31 generates the appearance model 35 will now be described in more detail with reference to Figures 4 to 8.

15 Figure 4 is a flow diagram illustrating the general processing steps performed by the appearance model generation unit 31 to generate the appearance model 35. As shown, there are three general steps S1, S3 and S5. 20 In step S1, a shape model is generated which models the variability of the head shapes within the training images. In step S3, a grey level model is generated which models the variability of the grey level of the heads in the training images. Finally, in step S5, the 25 shape model and the grey level model are used to generate an appearance model which collectively models the way in which both the shape and the grey level varies within the heads in the training images.

30 Figure 5 is a flow diagram illustrating the steps involved in generating the shape model in step S1 of Figure 4. As shown, in step S11, landmark points are placed on the heads in the training images (the selected frames from the video sequence and the target image) 35 manually by the user via the user interface 33. In

particular, in step S11, each training image is displayed in turn on the display 11 and the user places the landmark points over the head. In this embodiment, 86 landmark points are placed over each head in order to delineate the main features in the head, e.g. the position of the hair line, neck, eyes, nose, ears and mouth. In order to be able to compare training faces, each landmark point is associated with the same point on each face. For example, landmark point LP<sub>8</sub> is associated with the bottom of the nose and landmark point LP<sub>6</sub> is associated with the left-hand corner of the mouth. Figure 6 shows an example of one of the training images with the landmark points positioned over the head and the table below identifies each landmark point with its associated position on the head.

Landmark Point	Associated Position	Landmark Point	Associated Position
LP <sub>1</sub>	Left corner of left eye	LP <sub>4</sub>	Eye, bottom
LP <sub>2</sub>	Right corner of right eye	LP <sub>45</sub>	Eye, top
LP <sub>3</sub>	Chin, bottom	LP <sub>46</sub>	Eye, bottom
LP <sub>4</sub>	Right corner of left eye	LP <sub>47</sub>	Eyebrow, lower
LP <sub>5</sub>	Left corner of right eye	LP <sub>48</sub>	Eyebrow, upper
LP <sub>6</sub>	Mouth, left	LP <sub>49</sub>	Cheek, left
LP <sub>7</sub>	Mouth, right	LP <sub>50</sub>	Cheek, right
LP <sub>8</sub>	Nose, bottom	LP <sub>51</sub>	Eyebrow, lower
LP <sub>9</sub>	Nose, between eyes	LP <sub>52</sub>	Eyebrow, upper
LP <sub>10</sub>	Upper lip, top	LP <sub>53</sub>	Eyebrow, lower
LP <sub>11</sub>	Lower lip, bottom	LP <sub>54</sub>	Eyebrow, upper
LP <sub>12</sub>	Neck, left, top	LP <sub>55</sub>	Eyebrow, lower
LP <sub>13</sub>	Neck, right, top	LP <sub>56</sub>	Eyebrow, upper

	Landmark Point	Associated Position	Landmark Point	Associated Position
5	LP <sub>14</sub>	Face edge left, level with nose	LP <sub>57</sub>	Eyebrow, lower
	LP <sub>15</sub>	Face edge	LP <sub>58</sub>	Eyebrow, upper
	LP <sub>16</sub>	Face edge right, level with nose	LP <sub>59</sub>	Eyebrow, lower
	LP <sub>17</sub>	Face edge	LP <sub>60</sub>	Eyebrow, upper
	LP <sub>18</sub>	Top of head	LP <sub>61</sub>	Eyebrow, lower
	LP <sub>19</sub>	Hair edge	LP <sub>62</sub>	Lower lip, top
	LP <sub>20</sub>	Hair edge	LP <sub>63</sub>	Centre forehead
	LP <sub>21</sub>	Hair edge	LP <sub>64</sub>	Upper lip, top left
	LP <sub>22</sub>	Hair edge	LP <sub>65</sub>	Upper lip, top right
	LP <sub>23</sub>	Hair edge	LP <sub>66</sub>	Lower lip, bottom right
10	LP <sub>24</sub>	Hair edge	LP <sub>67</sub>	Lower lip, bottom left
	LP <sub>25</sub>	Hair edge	LP <sub>68</sub>	Eye, top left
	LP <sub>26</sub>	Hair edge	LP <sub>69</sub>	Eye, top right
	LP <sub>27</sub>	Hair edge	LP <sub>70</sub>	Eye, bottom right
	LP <sub>28</sub>	Hair edge	LP <sub>71</sub>	Eye, bottom left
15	LP <sub>29</sub>	Bottom, far left	LP <sub>72</sub>	Eye, top left
	LP <sub>30</sub>	Bottom, far right	LP <sub>73</sub>	Eye, top right
	LP <sub>31</sub>	Shoulder	LP <sub>74</sub>	Eye, bottom right
	LP <sub>32</sub>	Shoulder	LP <sub>75</sub>	Eye, bottom left
	LP <sub>33</sub>	Bottom, left	LP <sub>76</sub>	Lower lip, top left
20	LP <sub>34</sub>	Bottom, middle	LP <sub>77</sub>	Lower lip, top right
	LP <sub>35</sub>	Bottom, right	LP <sub>78</sub>	Chin, left
	LP <sub>36</sub>	Left forehead	LP <sub>79</sub>	Chin, right
	LP <sub>37</sub>	Right forehead	LP <sub>80</sub>	Neck, left
	LP <sub>38</sub>	Centre, between eyebrows	LP <sub>81</sub>	Neckline, left
25	LP <sub>39</sub>	Nose, left	LP <sub>82</sub>	Neckline
	LP <sub>40</sub>	Nose, right	LP <sub>83</sub>	Neckline, right

Landmark Point	Associated Position	Landmark Point	Associated Position
LP <sub>41</sub>	Nose edge, left	LP <sub>84</sub>	Neck, right
LP <sub>42</sub>	Nose edge, right	LP <sub>85</sub>	Hair edge
LP <sub>43</sub>	Eye, top	LP <sub>86</sub>	Hair edge

5

The result of this manual placement of the landmark points is a table of landmark points for each training image, which identifies the (x,y) coordinates of each landmark point within the image. The modelling technique used in this embodiment works by examining the statistics of these coordinates over the training set. In order to be able to compare equivalent points from different images, the heads must be aligned with respect to a common set of axes. This is achieved, in step S13, by iteratively rotating, scaling and translating the set of coordinates for each head so that they all approximately fill the same reference frame. The resulting set of coordinates for each head form a shape vector ( $x$ ) whose elements correspond to the coordinates of the landmark points within the reference frame. In other words, the shape and pose of each training head is represented by a vector ( $x$ ) of the following form:

$$x = [x_0, y_0, x_1, y_1, x_2, y_2, \dots, x_{85}, y_{85}]^T$$

This iterative alignment process is described in detail in the above paper by Cootes et al and will not be described in detail here. The shape model is then generated in step S15 by performing a principal component analysis (PCA) on the set of shape training vectors generated in step S13. An overview of this principal component analysis will now be given. (The reader is directed to a book by W.J. Krzanowski entitled

"Principles of Multivariate Analysis - A User's Perspective", 1998, (Oxford Statistical Science Series) for a more detailed discussion of principal component analysis.)

5

A principal component analysis of a set of training data finds all possible modes of variation within the training data. However, in this case, since the landmark points on the training heads do not move about independently, 10 i.e. their positions are partially correlated, most of the variation in the training faces can be explained by just a few modes of variation. In this embodiment, the main mode of variation between the training faces is likely to be the difference between the shape of the 15 first actor's head and the shape of the second actor's head. The other main modes of variation will describe the changes in shape and pose of the first actor's head within the selected source video frames. The principal component analysis of the shape training vectors  $x^i$  generates a shape model (matrix  $P_s$ ) which relates each 20 shape vector to a corresponding vector of shape parameters, by:

$$b_s^i = P_s(x^i - \bar{x}) \quad (1)$$

where  $x^i$  is a shape vector,  $\bar{x}$  is the mean shape vector 25 from the shape training vectors and  $b_s^i$  is a vector of shape parameters for the shape vector  $x^i$ . The matrix  $P_s$  describes the main modes of variation of the shape and pose within the training heads; and the vector of shape 30 parameters ( $b_s^i$ ) for a given input head has a parameter associated with each mode of variation whose value relates the shape of the given input head to the corresponding mode of variation. For example, if the heads in the training images include thin heads, normal width heads and broad heads, then one mode of variation

which will be described by the shape model ( $P_s$ ) will have an associated parameter within the vector of shape parameters ( $b_s$ ) which affects, amongst other things, the width of an input head. In particular, this parameter 5 might vary from -1 to +1, with parameter values near -1 being associated with thin heads, with parameter values around 0 being associated with normal width heads and with parameter values near +1 being associated with broad heads.

10 Therefore, the more modes of variation which are required to explain the variation within the training data, the more shape parameters are required within the shape parameter vector  $b_s^i$ . In this embodiment, for the 15 particular training images used, 20 different modes of variation of the shape and pose must be modelled in order to explain 98% of the variation which is observed within the training heads. Therefore, using the shape model ( $P_s$ ), the shape and pose of each head within the training 20 images can be approximated by just 20 shape parameters. As those skilled in the art will appreciate, in other embodiments, more or less modes of variation may be required to achieve the same model accuracy. For example, if the first actor's head does not move or 25 change shape significantly during the video sequence, then fewer modes of variation are likely to be required for the same accuracy.

30 In addition to being able to determine a set of shape parameters  $b_s^i$  for a given shape vector  $x^i$ , equation 1 can be solved with respect to  $x^i$  to give:

$$x^i = \bar{x} + P_s^T b_s^i \quad (2)$$

since  $P_s P_s^T$  equals the identity matrix. Therefore, by

modifying the set of shape parameters ( $b_s$ ), within suitable limits, new head shapes can be generated which will be similar to those in the training set.

5 Once the shape model has been generated, a similar model is generated to model the grey level within the training heads. Figure 7 illustrates the processing steps involved in generating this grey level model. As shown, in step S21, each training head is deformed to the mean shape. This is achieved by warping each head until the corresponding landmark points coincide with the mean landmark points (obtained from  $\bar{x}$ ) depicting the shape and pose of the mean head. Various triangulation techniques can be used to deform each training head to the mean shape. The preferred way, however, is based on a technique developed by Bookstein based on thin plate splines, as described in "Principle Warps: Thin-Plate Splines and the Decomposition of Deformations" IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 11, No. 6, pp 567-585, 1989, the contents of which are incorporated herein by reference.

10

15

20

In step S23, a grey level vector ( $g^i$ ) is determined for each shape-normalised training head, by sampling the grey level value at 10,656 evenly distributed points over the shape-normalised head. A principal component analysis of these grey level vectors is then performed in step S25. As with the principal component analysis of the shape training vectors, the principal component analysis of the grey level vectors generates a grey level model (matrix  $P_g$ ) which relates each grey level vector to a corresponding vector of shape parameters, by:

25

30

$$b_g^i = P_g(g^i - \bar{g}) \quad (3)$$

where  $g^i$  is a grey level vector,  $\bar{g}$  is the mean grey level

vector from the grey level training vectors and  $b_g^i$  is a vector of grey level parameters for the grey level vector  $g^i$ . The matrix  $P_g$  describes the main modes of variation of the grey level within the shape-normalised training heads. In this embodiment, 30 different modes of variation of the grey level must be modelled in order to explain 98% of the variation which is observed within the shape-normalised training heads. Therefore, using the grey level model ( $P_g$ ), the grey level of each shape-normalised training head can be approximated by just 30 grey level parameters.

In the same way that equation 1 was solved with respect to  $x^i$ , equation 3 can be solved with respect to  $g^i$  to give:

$$g^i = \bar{g} + P_g^T b_g^i \quad (4)$$

since  $P_g P_g^T$  equals the identity matrix. Therefore, by modifying the set of grey level parameters ( $b_g$ ), within suitable limits, new shape-normalised grey level faces can be generated which will be similar to those in the training set.

As mentioned above, the shape model and the grey level model are used to generate an appearance model which collectively models the way in which both the shape and the grey level varies within the heads of the training images. A combined appearance model is generated because there are correlations between the shape and grey level variations, which can be used to reduce the number of parameters required to describe the total variation within the training faces by performing a further principal component analysis on the shape and grey level parameters. Figure 8 shows the processing steps involved

in generating the appearance model using the shape and grey level models previously determined. As shown, in step S31, shape parameters ( $b_s^i$ ) and grey level parameters ( $b_g^i$ ) are determined for each training head from equations 1 and 3 respectively. The resulting parameters are concatenated and a principal component analysis is performed on the concatenated vectors to determine the appearance model (matrix  $P_{sg}$ ) such that:

$$c^i = P_{sg} b_{sg} \quad (5)$$

10

where  $c^i$  is a vector of appearance parameters controlling both the shape and grey levels and  $b_{sg}$  are the concatenated shape and grey level parameters. In this embodiment, 40 different modes of variation and hence 40 appearance parameters are necessary to model 98% of the variation found in the concatenated shape and grey level parameters. As those skilled in the art will appreciate, this represents a considerable compression over the 86 landmark points and the 10,656 grey level values originally used to describe each head.

15

#### HEAD REGENERATION

In addition to being able to represent an input head by the 40 appearance parameters ( $c$ ), it is also possible to use those appearance parameters to regenerate the input head. In particular, by combining equation 5 with equations 1 and 3 above, expressions for the shape vector ( $x^i$ ) and for the grey level vector ( $g^i$ ) can be determined as follows:

20

$$x^i = \bar{x} + Q_s c \quad (6)$$

$$g^i = \bar{g} + Q_g c \quad (7)$$

where  $Q_s$  is obtained from  $P_{sg}$  and  $P_s$ , and  $Q_g$  is obtained from  $P_{sg}$  and  $P_g$  (and where  $Q_s$  and  $Q_g$  map the value of  $c$  to changes in the shape and shape normalised grey level data). However, in order to regenerate the head, the  
5 shape-free grey level image generated from the vector  $g^i$  must be warped to take into account the shape of the head as described by the shape vector  $x^i$ . The way in which this warping of the shape-free grey level image is performed will now be described.

10

When the shape-free grey level vector ( $g^i$ ) was determined in step S23, the grey level at 10,656 points over the shape-free head was determined. Since each head is deformed to the same mean shape, these 10,656 points are  
15 extracted from the same position within each shape-normalised training head. If the position of each of these points is determined in terms of the positions of three landmark points, then the corresponding position of that point in a given face can be determined from the position of the corresponding three landmark points in  
20 the given face (which can be found from the generated shape vector  $x^i$ ). In this embodiment, a mesh of triangles is defined which overlays the landmark points such that the corners of each triangle corresponds to one  
25 of the landmark points. Figure 9 shows the head shown in Figure 6 with the mesh of triangles placed over the head in accordance with the positions of the landmark points.

Figure 10 shows a single point  $p$  located within the  
30 triangle formed by landmark points  $LP_i$ ,  $LP_j$  and  $LP_k$ . The position of point  $p$  relative to the origin (O) of the reference frame can be expressed in terms of the position of the landmark points  $LP_i$ ,  $LP_j$  and  $LP_k$ . In particular, the vector between the origin and the point  $p$  can be  
35 expressed by the following:

$$V_p = aP_i + bP_j + cP_k \quad (8)$$

where a, b and c are scalar values and  $P_i$ ,  $P_j$  and  $P_k$  are the vectors describing the positions of the landmark points  $LP_i$ ,  $LP_j$  and  $LP_k$ . In the shape-normalised heads, 5 the positions of the 10,656 points and the position of the landmark points LP are known, and therefore, the values of a, b and c for each of the 10,656 points can be determined. These values are stored and then used together with the positions of the corresponding landmark 10 points in the given face (determined from the generated shape vector  $x^i$ ) to warp the shape-normalised grey level head, thereby regenerating the head from the appearance parameters (c).

15 TARGET VIDEO SEQUENCE GENERATION

A description will now be given of the way in which the target video sequence is generated from the source video sequence. As shown in Figure 11, the source video sequence is input to a target video sequence generation 20 unit 51 which processes the source video sequence using a set of difference parameters 53 to generate and to output the target video sequence.

Figure 12 is a flow diagram illustrating the processing 25 steps involved in generating these difference parameters. As shown, in step S41, the appearance parameters ( $c_s$ ) for an example of the first actor's head (from one of the training images) and the appearance parameters ( $c_t$ ) for the second actor's head (from the target image) are 30 determined. This is achieved by determining the shape parameter vector ( $b_s$ ) and the grey level parameter vector ( $b_t$ ) for each of the two images and then calculating the corresponding appearance parameters by inserting these shape and grey level parameters into equation 5. In step

S43, a set of difference parameters are then generated by subtracting the appearance parameters ( $c_s$ ) for the first actor's head from the appearance parameters ( $c_T$ ) for the second actor's head, i.e. from:

5

$$C_{dif} = C_T - C_s \quad (9)$$

In order that these difference parameters only represent differences in the general shape and grey level of the two actors' heads, the pose and expression on the first actor's head in the training image used in step S41 should match, as closely as possible, the pose and expression of the second actor's head in the target image. Therefore, care has to be taken in selecting the source video frame used to calculate the appearance parameters in step S41.

10

The processing steps required to generate the target video sequence from the source video sequence will now be described in more detail with reference to Figure 13. As shown, in step S51, the appearance parameters ( $c_s^i$ ) for the first actor's head in the current video frame are automatically calculated. The way that this is achieved in this embodiment will be described later. In step S53, the difference parameters ( $C_{dif}$ ) are added to the appearance parameters for the current source head to generate:

15

20

$$C_{mod}^i = C_s^i + C_{dif} \quad (10)$$

25

The resulting appearance parameters ( $C_{mod}^i$ ) are then used in step S55 to regenerate the head for the current video frame. In particular, the shape vector ( $x^i$ ) and the shape-normalised grey level vector ( $g^i$ ) are generated from equations 6 and 7 using the modified appearance

parameters ( $c^i_{mod}$ ) and then the shape-normalised grey level image generated by the grey level vector ( $g^i$ ) is then warped using the 10,656 stored scalar values for a, b and c and the shape vector ( $x^i$ ), in the manner described above, to regenerate the head. In this embodiment, since the resolution of the video frame is 500 x 500 pixels interpolation is used to determine the grey level values for pixels located between the 10,656 points. The regenerated head is then composited, in step S57, into the source video frame to generate a corresponding target video frame. A check is then made, in step S59, to determine whether or not there are any more source video frames. If there are then the processing returns to step S51 where the procedure described above is repeated for the next source video frame. If there are no more source video frames, then the processing ends.

Figure 14 illustrates the results of this animation technique. In particular, Figure 14a shows three frames of the source video sequence, Figure 14b shows the target image (which in this embodiment is computer-generated) and Figure 14c shows the corresponding three frames of the target video sequence obtained in the manner described above. As can be seen, an animated sequence of the computer-generated character has been generated from a video clip of a real person and a single image of the computer-generated character.

30 AUTOMATIC GENERATION OF APPEARANCE PARAMETERS

In step S51, appearance parameters for the first actor's head in each video frame were automatically calculated. In this embodiment, this is achieved in a two-step process. In the first step, an initial set of appearance parameters for the head is found using a simple and rapid

technique. For all but the first frame of the source video sequence, this is achieved by simply using the appearance parameters ( $c_s^{i-1}$ ) from the preceding video frame (before modification in step S53). As described 5 above, the appearance parameters ( $c$ ) effectively define the shape and grey level of the head, but they do not define the scale, position and orientation of the head within the video frame. For all but the first frame in the source video sequence, these also can be initially 10 estimated to be the same as those for the head in the preceding frame.

For the first frame, if it is one of the training images input to the appearance model generation unit 31, then 15 the scale, position and orientation of the head within the frame will be known from the manual placement of the landmark points and the appearance parameters can be generated from the shape parameters and the shape-normalised grey level parameters obtained during 20 training. If the first frame is not one of the training images, as in the present embodiment, then the initial estimate of the appearance parameters is set to the mean set of appearance parameters (i.e. all the appearance 25 parameters are zero) and the scale, position and orientation is initially estimated by the user manually placing the mean face over the head in the first frame.

In the second step, an iterative technique is used in order to make fine adjustments to the initial estimate of 30 the appearance parameters. The adjustments are made in an attempt to minimise the difference between the head described by the appearance parameters (the model head) and the head in the current video frame (the image head). With 30 appearance parameters, this represents a 35 difficult optimisation problem. However, since each

attempt to match the model head to a new image head, is actually a similar optimisation problem, it is possible to learn in advance how the parameters should be changed for a given difference. For example, if the largest 5 differences between the model head and the image head occur at the sides of the head, then this implies that a parameter that adjusts the width of the model head should be adjusted.

10 In this embodiment, it is assumed that there is a linear relationship between the error ( $\delta c$ ) in the appearance parameters (i.e. the change to be made) and the difference ( $\delta I$ ) between the model head and the image head, i.e.

$$\delta c = A\delta I \quad (11)$$

15 In this embodiment, the relationship (A) was found by performing multiple multivariate linear regressions on a large sample of known model displacements ( $\delta c$ ) and the corresponding difference images ( $\delta I$ ). These large sets 20 of random displacements were obtained by perturbing the true model parameters for the images in the training set by a known amount. As well as perturbations in the model parameters, small displacements in the scale, position and orientation were also modelled and included in the 25 regression; for simplicity of notation, the parameters describing scale, position and orientation were regarded simply as extra elements within the vector  $\delta c$ . In this embodiment, during the training, the difference between the model head and the image head was determined from the 30 difference between the corresponding shape normalised grey level vectors. In particular, for the current location within the video frame, the actual shape-normalised grey level vector  $g^i$  was determined (in the

manner described above with reference to Figure 7) which was then compared with the grey level vector  $g^m$  obtained from the current appearance parameters using equation 7 above, i.e.

5

$$\delta I = \delta g = g^i - g^m \quad (12)$$

10

After A has been determined from this training phase, an iterative method for solving the optimisation problem can be determined by calculating the grey level difference vector,  $\delta g$ , for the current estimate of the appearance parameters and then generating a new estimate for the appearance parameters from:

$$c' = c - A\delta g \quad (13)$$

15

(noting here that the vector c includes the appearance parameters and the parameters defining the current estimate of the scale, position and orientation of the head within the image).

20

#### ALTERNATIVE EMBODIMENTS

As those skilled in the art will appreciate, a number of modifications can be made to the above embodiment. A number of these modifications will now be described.

25

In the above embodiment, the target image frame illustrated a computer generated head. This is not essential. For example, the target image might be a hand-drawn head or an image of a real person. Figures 14d and 14e illustrate how an embodiment with a hand-drawn character might be used in character animation. In particular, Figure 14d shows a hand-drawn sketch of a character which when combined with the frames from the source video sequence (some of which are shown in Figure 14a) generate a target video sequence, some frames of

30

which are shown in Figure 14e. As can be seen from a comparison of the corresponding frames in the source and target video frames, the hand-drawn sketch has been animated automatically using this technique. As those skilled in the art will appreciate, this is a much quicker and simpler technique for achieving computer animation, as compared with existing systems which require the animator to manually create each frame of the animation. In particular, in this embodiment, all that is required is a video sequence of a real life actor acting out the scene to be animated, together with a single sketch of the character to be animated.

In the above embodiments, the head, neck and shoulders of the first actor in the video sequence was modified using the corresponding head, neck and shoulders from the target image. This is not essential. As those skilled in the art will appreciate, only those parts of the image in and around the landmark points will be modified.

Therefore, if the landmark points are only placed in and around the first actor's face, then only the face in the video sequence will be modified. This animation technique can be applied to any part of the body which is deformable and even to other animals and objects. For example, the technique may be applied to just the lips in the video sequence. Such an embodiment could be used in film dubbing applications in order to synchronise the lip movements with the dubbed sound. This animation technique might also be used to give animals and other objects human-like characteristics by combining images of them with a video sequence of an actor.

In the above embodiment, 86 landmark points were placed around the head, neck and shoulders of the test images.

As those skilled in the art will appreciate, more or less landmark points. Similarly, the number of points in the shape-normalised head for which a grey level value is sampled also depends upon the required accuracy of the  
5 system.

In the above embodiment, the shape and grey level of the heads in the source video sequence and in the target image were modelled using principal component analysis.  
10 As those skilled in the art will appreciate, by modelling the features of the heads in this way, it is possible to accurately model each head by just a small number of parameters. However, other modelling techniques, such as vector quantisation and wavelet techniques can be used.  
15 Furthermore, it is not essential to model each of the heads, however, doing so results in fewer computations being required in order to modify each frame in the source video sequence. In an embodiment where no modelling is performed, the difference parameters could simply be the difference between the location of the  
20 landmark points in the target image and in the selected frame from the source video sequence. It may also include a set of different signals indicative of the difference between the grey level values from the  
25 corresponding heads.

In the above embodiment, the shape parameters and the grey level parameters were combined to generate the appearance parameters. This is not essential. A separate set of shape difference parameters and grey level difference parameters could be calculated however this is not preferred, since it increases the number of parameters which have to be automatically generated for each source video frame in step S51 described above.  
30  
35

In the above embodiments, the source video sequence and the target image were both black and white. The present invention can also be applied to colour images. In particular, if each pixel in the source video frames and in the target image has a corresponding red, green and blue pixel value, then instead of sampling the grey level at each of the 10,656 points in the shape-normalised head, the colour embodiment would sample each of the red, green and blue values at those points. The remaining processing steps would essentially be the same except that there would be a colour level model which would model the variations in the colour in the training images. Further, as those skilled in the art will appreciate, the way in which colour is represented in such an embodiment is not important. In particular, rather than each pixel having a red, green and blue value, they might be represented by a chrominance and a luminance component or by hue, saturation and value components. Both of these embodiments would be simpler than the red, green and blue embodiment, since the image search which is required during the automatic calculation of the appearance parameters in step S51 could be performed using only the luminance or value component. In contrast, in the red, green and blue colour embodiment, each of these terms would have to be considered in the image search.

In the above embodiment, during the automatic generation of the appearance parameters, and in particular during the iterative updating of these appearance parameters using equation 13 above, the grey level value at each of the 10,656 points within the grey level vector obtained for the current location within the video frame and within the corresponding grey level vector obtained from the model were considered at each iteration. In an

- alternative embodiment, the resolution employed at each iteration might be changed. For example, in the first iteration, the grey level value at 1000 points might be considered to generate the difference vector  $\delta g$ . Then,  
5 in the second iteration, the grey level value at 3000 points might be considered during the determination of the difference vector  $\delta g$ . Then for subsequent iterations the grey level value at each of the 10,656 points could be considered during the determination of the different  
10 vector  $\delta g$ . By performing the search at difference resolutions, the convergence of the automatically generated appearance parameters for the current head in the source video sequence can be achieved more quickly.
- 15 In the above embodiment, a single target image was used to modify the source video sequence. As those skilled in the art will appreciate, two or more images of the second actor could be used during the training of the appearance model and during the generation of the difference  
20 parameters. In such an embodiment, during the determination of the difference parameters, each of the target images would be paired with a similar image from the source video sequence and the difference parameters determined from each would be averaged to determine a set  
25 of average difference parameters.
- In the above embodiment, the difference parameters were determined by comparing the image of the first actor from one of the frames from the source video sequence with the  
30 image of the second actor in the target image. In an alternative embodiment, a separate image of the first actor may be provided which does not form part of the source video sequence.
- 35 In the above embodiments, each of the images in the

source video sequence and the target image were two-dimensional images. The above technique could be adapted to work with 3D modelling and animations. In such an embodiment, the training data would comprise a set of 3D models instead of 2D images. Instead of the shape model being a two-dimensional triangular mesh, it would be a three-dimensional triangular mesh. The 3D models in the training set would have to be based on the same standardised mesh, i.e., like the 2D embodiment, they would each have the same number of landmark points with each landmark point being in the same corresponding position in each model. The grey level model would be sampled from the texture image mapped onto the three-dimensional triangles formed by the mesh of landmark points. The three-dimensional models may be obtained using a three-dimensional scanner which typically work either by using laser range-finding over the object or by using one or more stereo pairs of cameras. The standardised 3D triangular mesh would then be fitted to the 3D model obtained from the scanner. Once a 3D appearance model has been created from the training models, new 3D models can be generated by adjusting the appearance parameters, and existing 3D models can be animated using the same differencing technique that was used in the two-dimensional embodiment described above.

In the above embodiment, the grey level vector was determined from the shape-normalised head of the first and second actors. Other types of grey level model might be used. For example, a profile of grey level values at each landmark point might be used instead of or in addition to the sampled grey level value across the object. The way in which such profiles might be generated and the way in which the appearance parameters would be automatically found during step S51 in such an

embodiment can be found in the above paper by Cootes et al and in the paper entitled "Automatic Interpretation and Coding of Face Images using Flexible Models" by Andreas Lanitis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, July 1997, the contents of which are incorporated herein by reference.

During training of the above embodiment, the landmark points were manually placed on each of the training images by the user. In an alternative embodiment, an existing model might be used to automatically locate the appearance parameters on the training faces. Depending on the result of this automatic placement of the landmark points, the user may have to manually adjust the position of some of the landmark points. However, even in this case, the automatic placement of the landmark points would considerably reduce the time required to train the system.

In the above embodiment, during the automatic determination of the appearance parameters for the first frame in the source video sequence, they were initially set to be equal to the mean appearance parameters and with the scale position and orientation set by the user.

In an alternative embodiment, an initial estimate of the appearance parameters and of the scale, position and orientation of the head within the first frame can be determined from the nearest frame which was a training image (which, in the first embodiment, was frame  $f^s$ ). However, this technique might not be accurate enough if the scale, position and/or orientation of the head has moved considerably between the first frame in the sequence and the first frame which was a training image. In this case, an initial estimate for the appearance parameters for the first frame can be the appearance

parameters corresponding to the training head which is the most similar to the head in the first frame (determined from a visual inspection), and an initial estimate of the scale, position and orientation of the  
5 head within the first frame can be determined by matching the head which can be regenerated from those appearance parameters against the first frame, for various scales, positions and orientations and choosing the scale, position and orientation which provides the best match.

.10

In the above embodiments, a set of difference parameters were identified which describe the main differences between the actor in the video sequence and the actor in the target image, which difference parameters were used  
15 to modify the video sequence so as to generate a target video sequence showing the second actor. In the embodiment, the set of difference parameters were added to a set of appearance parameters for the current frame being processed. In an alternative embodiment, the  
20 difference parameters may be weighted so that, for example, the target video sequence shows an actor having characteristics from both the first and second actors.

25

In the above embodiment, a target image was used to modify each frame within a video sequence of frames. In an alternative embodiment, the target image might be used  
30 to modify a single source image. In this case, the difference parameters might be weighted in the manner described above so that the resulting object in the image is a cross between the object in the source image and the object in the target image. Alternatively, two source images might be provided, with the difference parameters being calculated with respect to one of the source images which are then applied to the second source image in  
35 order to generate the desired target image.

CLAIMS:

1. An image processing apparatus comprising:
  - means for receiving a source image of a first object;
  - means for receiving a target image of a second object;
  - means for comparing an image of the first object with the image of the second object to generate a difference signal; and
  - means for modifying the source image of the first object using said difference signal to generate a target image having characteristics of the first and second objects.
- 15 2. An image processing apparatus comprising:
  - means for receiving a source animated sequence of frames showing a first object;
  - means for receiving a target image showing a second object;
  - means for comparing an image of the first object with the image of the second object to generate a difference signal; and
  - means for modifying the image of the first object in each frame of said sequence of frames using said difference signal to generate a target animated sequence of frames showing the second object.
- 30 3. An apparatus according to claim 2, wherein said first object moves within said animated sequence of frames, and wherein said modifying means is arranged so that the target animated sequence of frames shows the second object moving in a similar manner.
- 35 4. An apparatus according to claim 2 or 3, wherein said

first object deforms over the sequence of frames and wherein said modifying means is arranged so that the target animated sequence of frames shows the second object deforming in a similar manner.

5

5. An apparatus according to claim 4, wherein said modifying means is operable for adding said difference signal to the image of said first object in each frame of said source animated sequence of frames to generate said target animated sequence of frames.

10

6. An apparatus according to any preceding claim, wherein said comparing means is operable to compare a first set of signals characteristic of the image of the first object with a second set of signals characteristic of the image of the second object to generate a set of difference signals.

15

7. An apparatus according to claim 6, wherein said modifying means is operable to use said set of difference signals to generate said target animated sequence of frames.

20

8. An apparatus according to claim 6 or 7, comprising processing means for processing the image of the second object and the image of the first object in order to generate said first and second sets of signals.

25

9. An apparatus according to claim 8, further comprising model means for modelling the visual characteristics of the first and second objects, and wherein said processing means is arranged to generate said first and second sets of signals using said model means.

30

35

10. An apparatus according to claim 9, wherein said model means is operable for modelling the variation of the appearance of the first and second objects within the received frames of the source animated sequence of frames and the received target image.
- 5
11. An apparatus according to claim 9 or 10, wherein said modifying means is operable (i) for determining, for the current frame being modified, a set of signals characteristic of the appearance of the first object in the frame using said model; (ii) to combine said set of signals with said difference signal to generate a set of modified signals; and (iii) to regenerate a corresponding frame using the modified set of signals and the model.
- 10
12. An apparatus according to any of claims 9 to 11, wherein said model means is operable for modelling the shape and colour of said first and second objects in said images.
- 15
13. An apparatus according to claim 12, wherein said model means is operable for modelling the shape and grey level of said first and second objects in said images.
- 20
14. An apparatus according to claim 12 or 13, comprising normalisation means for normalising the shape of said first and second objects in said images and wherein said model means is operable for modelling the colour within the shape-normalised first and second objects.
- 25
15. An apparatus according to any of claims 9 to 14, further comprising training means, responsive to the identification of the location of a plurality of points over the first and second objects in a set of training images, for training said model means to model the
- 30
- 35

variation of the position of said points within said set of training images.

5        16. An apparatus according to any of claims 9 to 15, wherein said training images include frames from the source animated sequence of frames and the target image.

10      17. An apparatus according to claim 14, 15 or 16, wherein said training means is operable to perform a principal component analysis modelling technique on the set of training images for training said model means.

15      18. An apparatus according to claim 17, wherein said training means is operable to perform a principal component analysis on a set of training data indicative of the shape of the objects within the training images for training said model means.

20      19. An apparatus according to claim 17 or 18, wherein said training means is operable to perform a principal component analysis on a set of data describing the colour over the objects within the training images for training said model means.

25      20. An apparatus according to claim 19 when dependent upon claim 18, wherein said training means is operable to perform a principal component analysis on a set of data obtained using a model obtained from the principal component analysis of the shape and the colour of the objects in the training images in order to train said model means to model both shape and colour variation within the objects of the training images.

30      21. An apparatus according to any of claims 6 to 20, wherein said comparing means is operable to subtract the

first set of signals characteristic of the image of the first object from the second set of signals characteristic of the image of the second object in order to generate said set of difference signals.

5

22. An apparatus according to any preceding claim, wherein said modifying means comprises means for processing each frame of the source animated sequence of frames in order to generate a set of signals characteristic of the first object in the frame being processed and wherein said modifying means is operable to modify the set of signals for the current frame being processed by combining them with said difference signal.

10

23. An apparatus according to any preceding claim, wherein said modifying means is arranged to modify each frame within the source animated sequence of frames in turn, in accordance with the position of the frame within the sequence of frames.

15

24. An apparatus according to any preceding claim, wherein said modifying means is arranged to automatically generate said target animated sequence from said source animated sequence and said difference signal.

20

25. An apparatus according to any preceding claim, wherein said image of the first object is obtained from a frame of said source animated sequence.

25

30 26. An apparatus according to any preceding claim, wherein said comparing means is arranged to compare a plurality of images of said first object with a plurality of images of said second object in order to generate a corresponding plurality of difference signals which are combined to generate said difference signal.

35

27. An apparatus according to claim 26, wherein said difference signal represents the average of said plurality of difference signals.

5 28. An apparatus according to any preceding claim, wherein the image of said first object is selected so as to generate a minimum difference signal.

10 29. An apparatus according to any preceding claim, wherein at least one of said first and second objects comprises a face.

15 30. An apparatus according to any preceding claim, wherein said target image comprises an image of a hand-drawn or a computer generated face.

31. A graphics processing apparatus comprising:  
means for receiving a source animated sequence of graphics data of a first object;  
20 means for receiving a target set of graphics data of a second object;  
means for comparing graphics data of the first object with graphics data of the second object to generate a difference signal; and  
25 means for modifying the graphics data in the animated sequence of graphics data using said difference signal to generate a target animated sequence of graphics data of the second object.

30 32. An apparatus according to claim 31, wherein said graphics data represents a 3D model or a 2D image.

33. A graphics processing apparatus comprising:  
means for receiving a source animated sequence of 3D models of a first object;

means for receiving a target 3D model of a second object;

5 means for comparing a 3D model of the first object with 3D the model of the second object to generate a difference signal; and

10 means for modifying each 3D model in the sequence 3D of models for the first object using said difference signal to generate a target animated sequence of 3D models for the second object.

10

34. An image processing apparatus comprising:

means for receiving a source sequence of frames recording a first animated object;

15

means for receiving a target image recording a second object;

means for comparing an image of the first object with the image of the second object to generate a set of difference signals; and

20

means for modifying the image of the first object in each frame of said sequence of frames using said set of difference signals to generate a target sequence of frames recording the second object animated in a similar manner to the animation of the first object.

25

35. An image processing apparatus comprising:

means for receiving a source sequence of frames showing a first object which deforms over the sequence of frames;

30

means for receiving a target image showing a second object;

means for comparing an image of the first object with the image of the second object to generate a difference signal; and

35

means for modifying the image of the first object in each frame of said sequence of frames using said

difference signal to generate a target sequence of frames showing the second object deforming in accordance with the deformations of the first object.

5       36. An image processing apparatus comprising:  
          means for receiving a source sequence of images comprising a first object which deforms over the sequence of images;

10.      means for receiving a target image comprising a second object;

          means for comparing the second object in the target image with the first object in a selected one of said images from said sequence of images and for outputting a comparison result;

15.      means for modifying the first object in each image of said source sequence of images using said comparison result to generate a target sequence of images comprising said second object which deforms in a similar manner to the way in which said first object deforms in said source sequence of images.

20.      37. An apparatus for performing computer animation, comprising:

25.      means for receiving signals representative of a film of a person acting out a scene;

          means for receiving signals representative of a character to be animated;

30.      means for comparing signals indicative of the appearance of the person with signals indicative of the appearance of the character to generate a difference signal; and

35.      means for modifying the signals representative of the film using said difference signal to generate modified signals representative of an animated film of the character acting out said scene.

38. An image processing method comprising the steps of:  
receiving a source animated sequence of frames  
showing a first object;  
receiving a target image showing a second object;  
5 comparing an image of the first object with the  
image of the second object to generate a difference  
signal; and  
modifying the image of the first object in each  
frame of said sequence of frames using said difference  
10 signal to generate a target animated sequence of frames  
showing the second object.
39. A method according to claim 38, wherein said first  
object moves within said animated sequence of frames, and  
15 wherein said modifying step is such that the target  
animated sequence of frames shows the second object  
moving in a similar manner.
40. A method according to claim 38 or 39, wherein said  
first object deforms over the sequence of frames and  
wherein said modifying step is such that the target  
animated sequence of frames shows the second object  
deforming in a similar manner.  
20
41. A method according to any of claims 38 to 40,  
wherein said modifying step combines said difference  
signal with the image of said first object in each frame  
of said source animated sequence of frames to generate  
said target animated sequence of frames.  
25
42. A method according to claim 41, wherein said  
modifying step adds said difference signal to the image  
of said first object in each frame of said source  
animated sequence of frames to generate said target  
animated sequence of frames.  
30
- 35

43. A method according to any of claims 38 to 42, wherein said comparing step compares a first set of signals characteristic of the image of the first object with a second set of signals characteristic of the image of the second object to generate a set of difference signals.
- 5
44. A method according to claim 43, wherein said modifying step uses said set of difference signals to generate said target animated sequence of frames.
- 10
45. A method according to claim 43 or 44, comprising the step of processing the image of the second object and the image of the first object in order to generate said first and second sets of signals.
- 15
46. A method according to claim 45, further comprising the step of modelling the visual characteristics of the first and second objects, and wherein said processing step generates said first and second sets of signals using the model generated by said modelling step.
- 20
47. A method according to claim 46, wherein said modelling step generates a model which models the variation of the appearance of the first and second objects within the received frames of the source animated sequence of frames and the received target image.
- 25
48. A method according to claim 46 or 47, wherein said modifying step (i) determines, for the current frame being modified, a set of signals characteristic of the appearance of the first object in the frame using said model; (ii) combines said set of signals with said difference signal to generate a set of modified signals; and (iii) regenerates a corresponding frame using the
- 30
- 35

modified set of signals and the model.

49. A method according to any of claims 46 to 48,  
wherein said modelling step generates a model which  
5 models the shape and colour of said first and second  
objects in said images.

50. A method according to claim 49, wherein said  
modelling step generates a model which models the shape  
10 and grey level of said first and second objects in said  
images.

51. A method according to claim 49 or 50, comprising the  
step of normalising the shape of said first and second  
15 objects in said images and wherein said modelling step  
generates a model which models the colour within the  
shape-normalised first and second objects.

52. A method according to any of claims 46 to 51,  
20 further comprising the steps of (i) identifying the  
location of a plurality of points over the first and  
second objects in a set of training images; (ii) and  
training said model to model the variation of the  
position of said points within said set of training  
25 images.

53. A method according to any of claims 46 to 52,  
wherein said training images include frames from the  
source animated sequence of frames and the target image.  
30

54. A method according to claim 51, 52 or 53, wherein  
said training step performs a principal component  
analysis modelling technique on the set of training  
images to train said model.  
35

55. A method according to claim 54, wherein said training step performs a principal component analysis on a set of training data indicative of the shape of the objects within the training images to train said model.

5

56. A method according to claim 54 or 55, wherein said training step performs a principal component analysis on a set of data describing the colour over the objects within the training images to train said model.

10

57. A method according to claim 56 when dependent upon claim 55, wherein said training step performs a principal component analysis on a set of data obtained using the models obtained from the principal component analysis of the shape and the colour of the objects in the training images in order to train said model to model both shape and colour variation within the objects of the training images.

15

58. A method according to any of claims 43 to 57, wherein said comparing step subtracts the first set of signals characteristic of the image of the first object from the second set of signals characteristic of the image of the second object in order to generate said set of difference signals.

25

59. A method according to any of claims 38 to 58, wherein said modifying step comprises the step of processing each frame of the source animated sequence of frames in order to generate a set of signals characteristic of the first object in the frame being processed and wherein said modifying step modifies the set of signals for the current frame being processed by combining them with said difference signal.

30

35

60. A method according to any of claims 38 to 59,  
wherein said modifying step is arranged to modify each  
frame within the source animated sequence of frames in  
turn, in accordance with the position of the frame within  
5 the sequence of frames.

61. A method according to any of claims 38 to 60,  
wherein said modifying step automatically generates said  
target animated sequence from said source animated  
10 sequence and said difference signal.

62. A method according to any of claims 38 to 61,  
wherein said image of the first object is obtained from  
a frame of said source animated sequence.

15 63. A method according to any of claims 38 to 62,  
wherein said comparing step compares a plurality of  
images of said first object with a plurality of images of  
said second object in order to generate a corresponding  
plurality of difference signals which are combined to  
20 generate said difference signal.

25 64. A method according to claim 63, wherein said  
difference signal represents the average of said  
plurality of difference signals.

65. A method according to any of claims 38 to 64,  
wherein the image of said first object is selected so as  
to generate a minimum difference signal.

30 66. A method according to any of claims 38 to 65,  
wherein at least one of said first and second objects  
comprises a face.

35 67. A method according to any of claims 38 to 66,

wherein said target image comprises an image of a hand-drawn or a computer generated face.

68. A graphics processing method comprising the steps  
5 of:

inputting a source animated sequence of graphics data for a first object;

10 comparing graphics data for the first object with graphics data for a second object to generate a difference signal; and

modifying the graphics data in the animated sequence of graphics data using said difference signal to generate a target animated sequence of graphics data for the second object.

15

69. A method according to claim 68, wherein said graphics data represents a 3D model or a 2D image.

20

70. A graphics processing method comprising the steps of:

receiving a source animated sequence of 3D models of a first object;

receiving a target 3D model of a second object;

25 comparing a 3D model of the first object with 3D the model of the second object to generate a difference signal; and

30 modifying each 3D model in the sequence 3D of models for the first object using said difference signal to generate a target animated sequence of 3D models for the second object.

35

71. An image processing method comprising the steps of:

receiving a source sequence of frames showing a first animated object;

receiving a target image showing a second object;

comparing an image of the first object with the image of the second object to generate a set of difference signals; and

5 modifying the image of the first object in each frame of said sequence of frames using said set of difference signals to generate a target sequence of frames showing the second object animated in a similar manner to the animation of the first object.

- 10 72. An image processing method comprising the steps of:  
receiving a source sequence of frames showing a first object which deforms over the sequence of frames;  
receiving a target image showing a second object;  
comparing an image of the first object with the  
15 image of the second object to generate a difference signal; and  
modifying the image of the first object in each frame of said sequence of frames using said difference signal to generate a target sequence of frames showing the second object deforming in accordance with the deformations of the first object.

- 20 73. An image processing method comprising the steps of:  
receiving a source sequence of images comprising a first object which deforms over the sequence of images;  
receiving a target image comprising a second object;  
comparing the second object in the target image with  
the first object in a selected one of said images from  
said sequence of images and for outputting a comparison  
25 result;  
modifying the first object in each image of said source sequence of images using said comparison result to generate a target sequence of images comprising said second object which deforms in a similar manner to the way in which said first object deforms in said source

sequence of images.

74. A computer animation method, comprising the steps of:

5 receiving signals representative of a film of a person acting out a scene;

receiving signals representative of a character to be animated;

10 comparing signals indicative of the appearance of the person with signals indicative of the appearance of the character to generate a difference signal; and

15 modifying the signals representative of the film using said difference signal to generate modified signals representative of an animated film of the character acting out said scene.

75. An apparatus according to any of claims 1 to 37, wherein said modifying means is operable to apply a weighting to said difference signal and to generate said target image using said weighted difference signal.

76. A storage medium storing processor implementable instructions for controlling a processor to carry out the method of any one of claims 38 to 74.

25 77. An electromagnetic or acoustic signal carrying processor implementable instructions for controlling a processor to carry out the method of any one of claims 38 to 74.

30 78. Processor implementable instructions for controlling a processor to carry out the method of any one of claims 38 to 74.

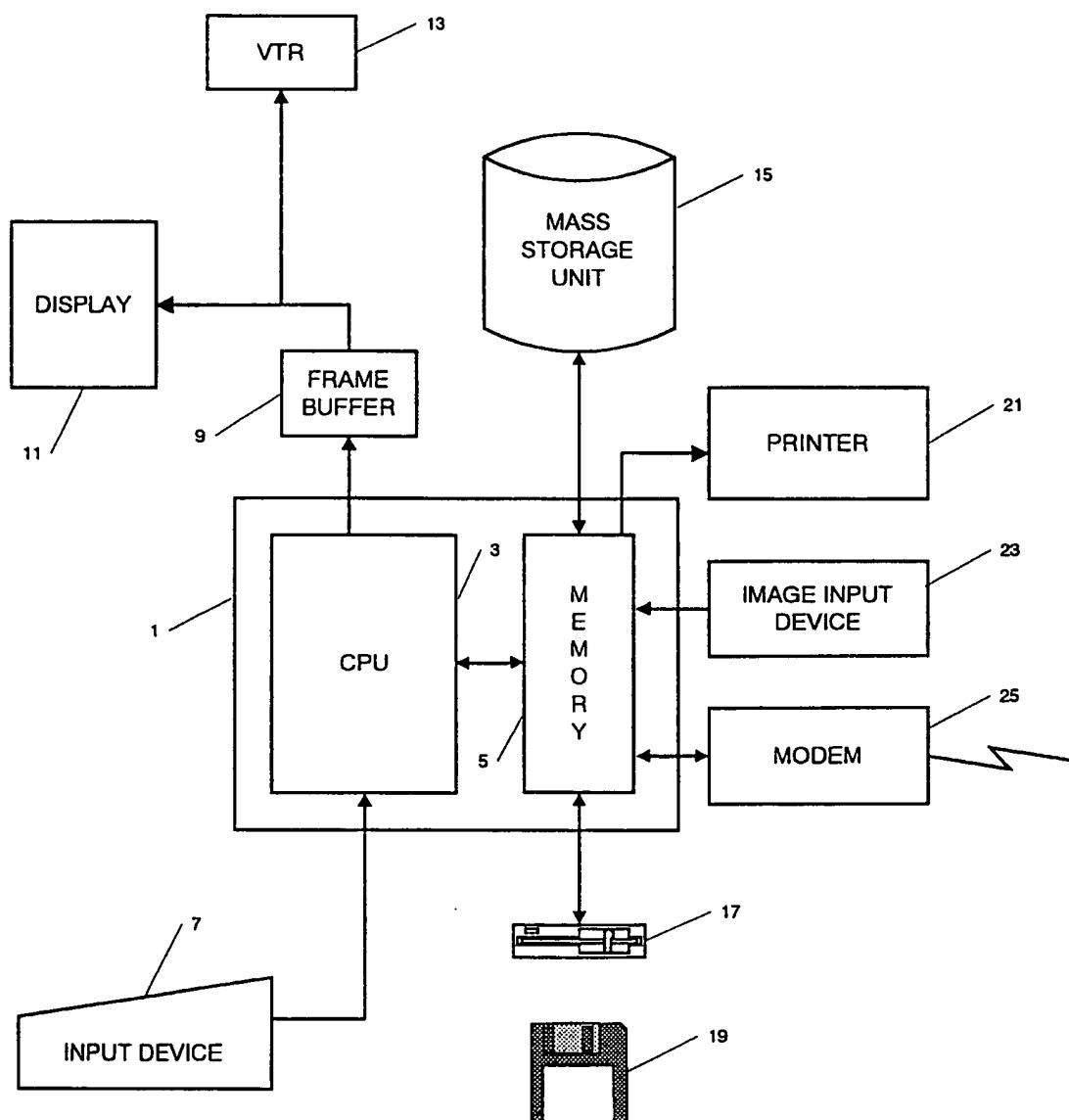


Fig. 1

2/13

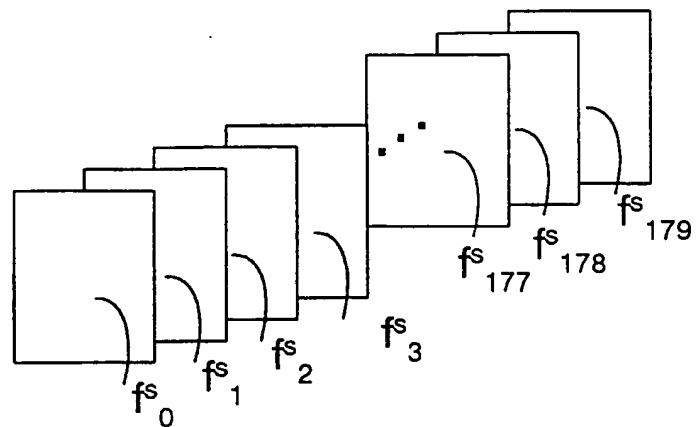


Fig. 2a

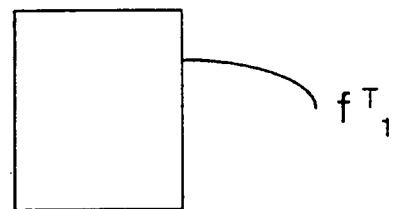


Fig. 2b

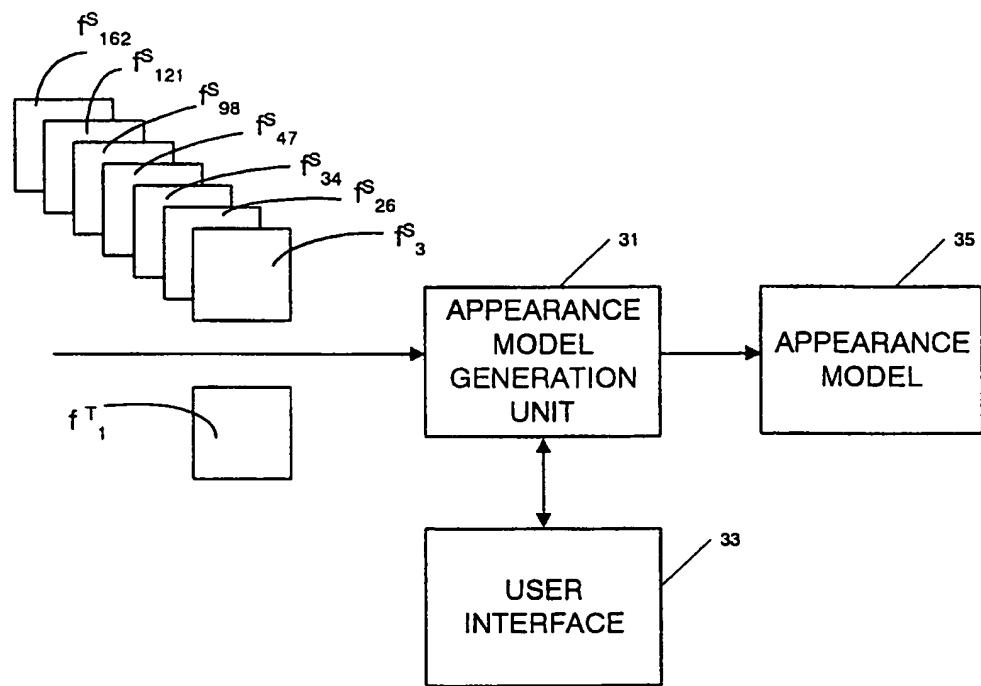


Fig. 3

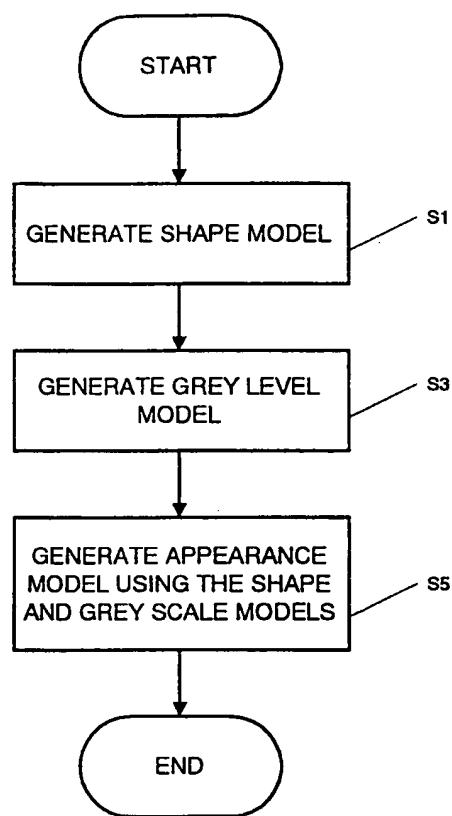


Fig. 4

5/13

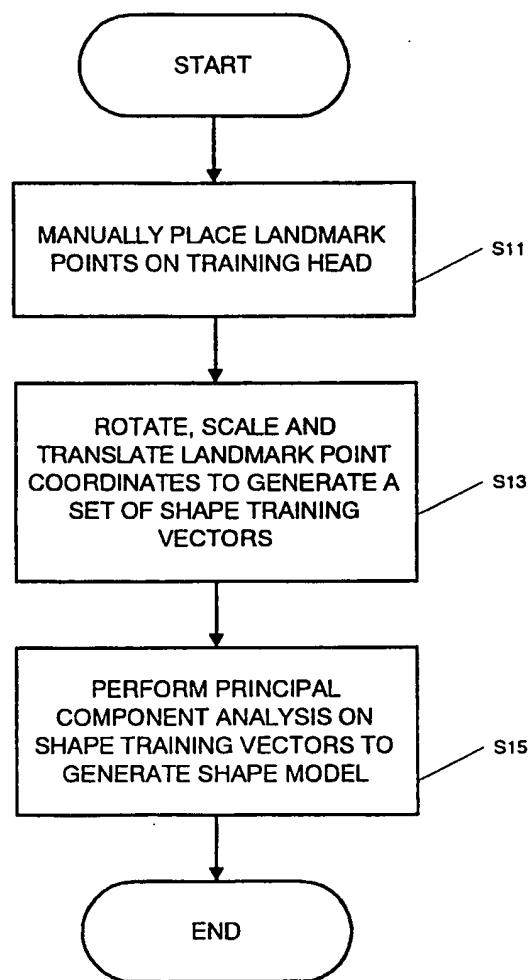


Fig. 5

6/13



Fig. 6

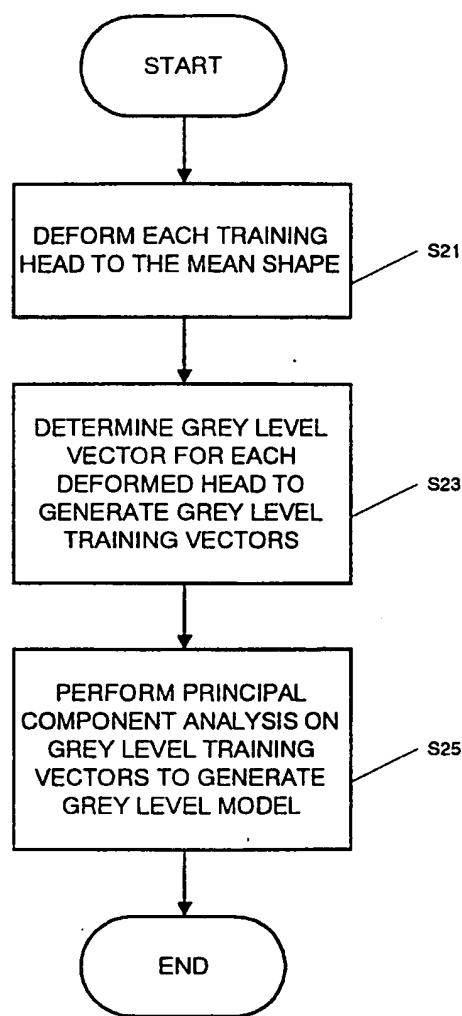


Fig. 7

8/13

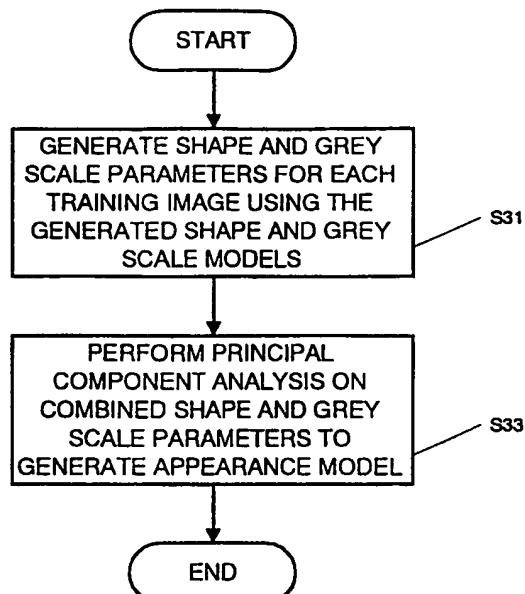


Fig. 8

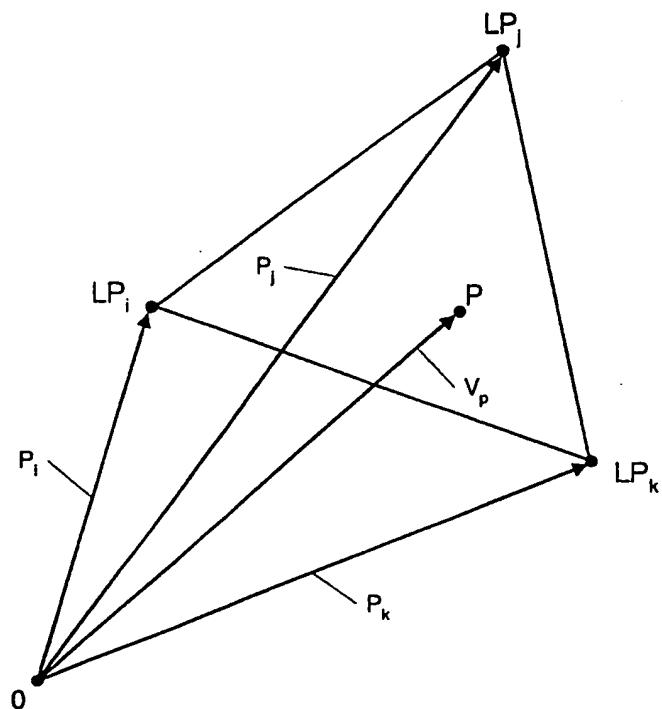


Fig. 10

9/13

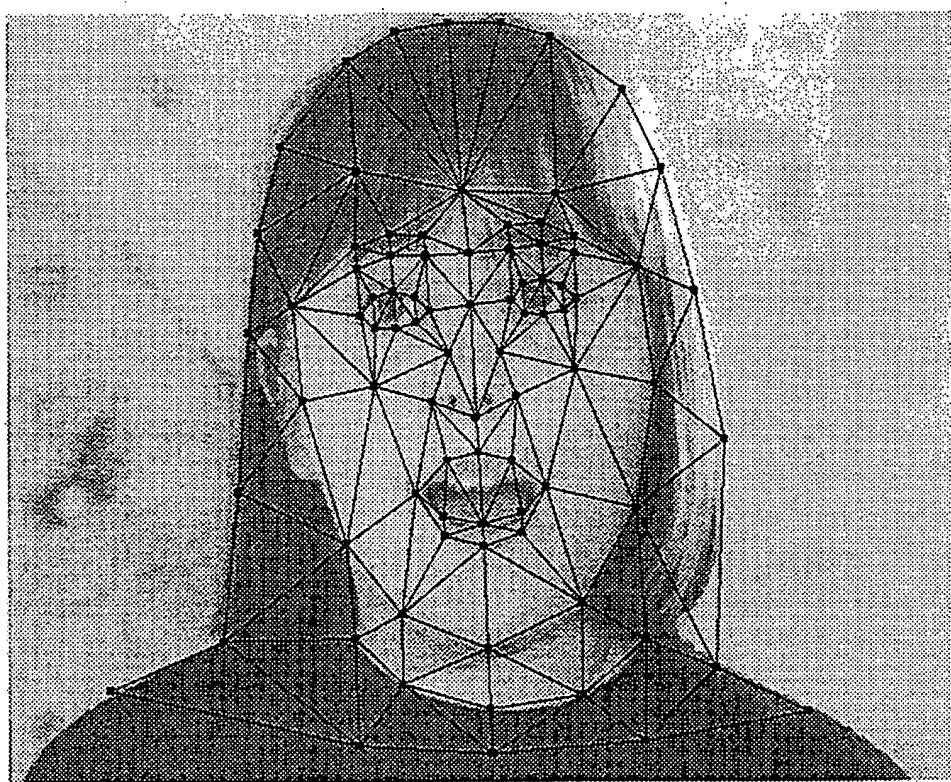


Fig. 9

10/13

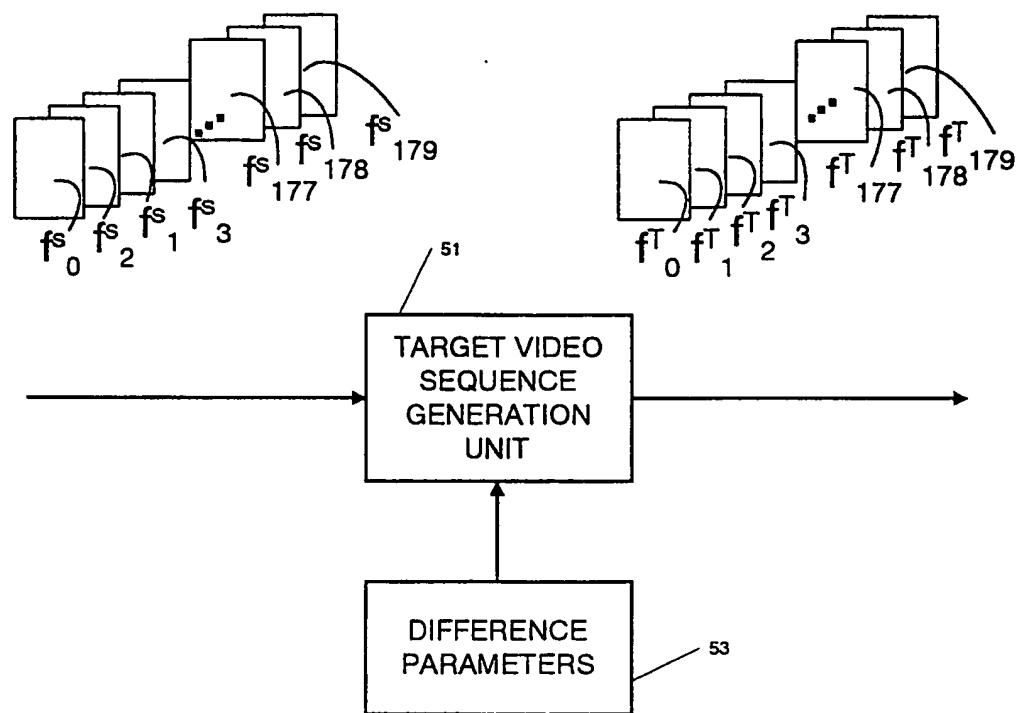


Fig. 11

11/13

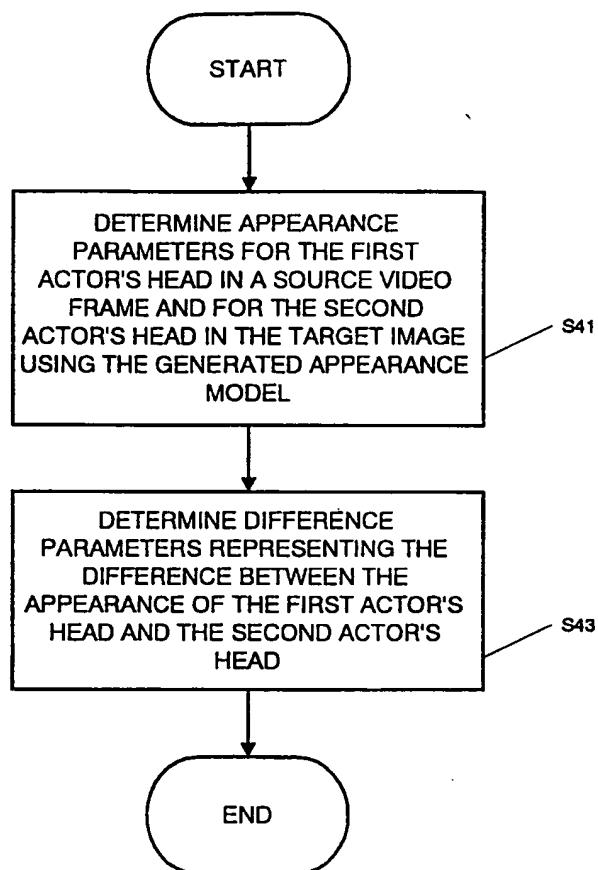


Fig. 12

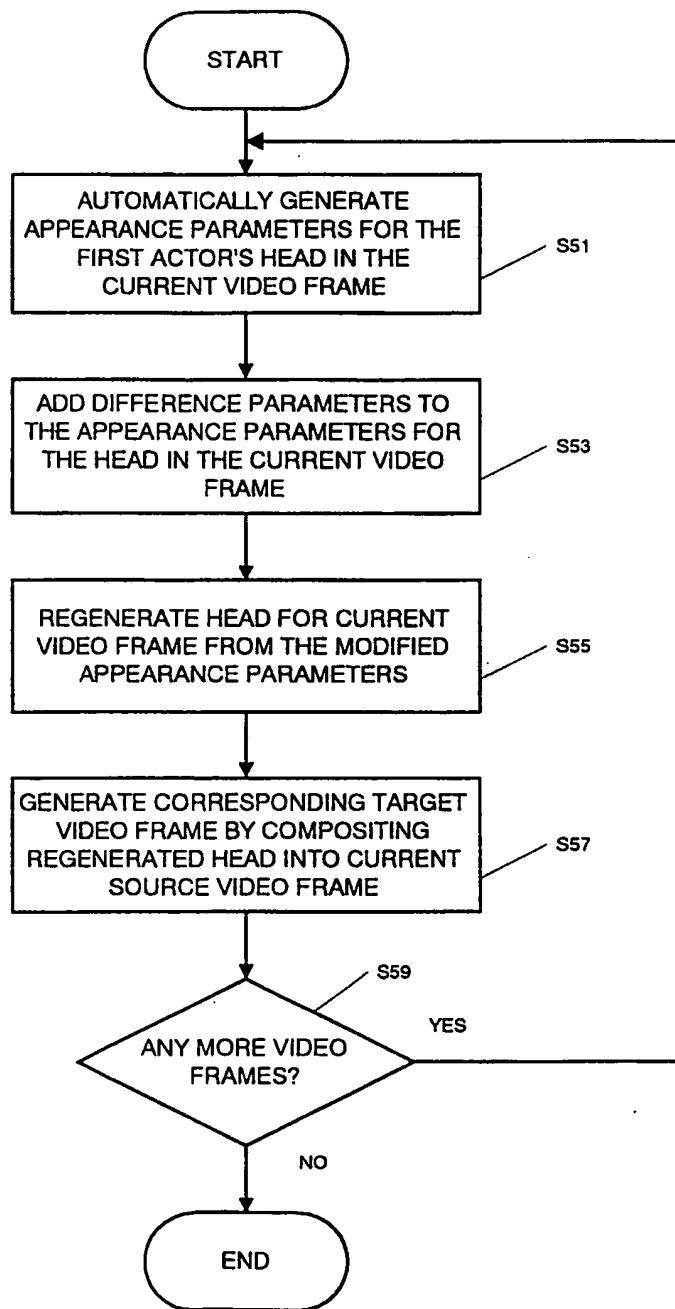


Fig. 13



Fig. 14a



Fig. 14b



Fig. 14c



Fig. 14d



Fig. 14e

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 99/03161

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 G06T15/70

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>NESI P ET AL: "TRACKING AND SYNTHESIZING FACIAL MOTIONS WITH DYNAMIC CONTOURS" REAL-TIME IMAGING, GB, ACADEMIC PRESS LIMITED, vol. 2, no. 2, April 1996 (1996-04), page 67-79 XP000656140 ISSN: 1077-2014 page 68, left-hand column, line 11 -page 69, left-hand column, line 2 ---</p> <p style="text-align: center;">-/-</p>	1-20, 22-25, 29, 31-62, 66-74, 76-78

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "Z" document member of the same patent family

Date of the actual completion of the international search

17 January 2000

Date of mailing of the international search report

25/01/2000

Name and mailing address of the ISA  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3018

Authorized officer

Burgaud, C

## INTERNATIONAL SEARCH REPORT

Inte onal Application No  
PCT/GB 99/03161

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 664 526 A (EASTMAN KODAK CO) 26 July 1995 (1995-07-26)  column 2, line 32 - line 38 column 4, line 6 - line 18 column 7, line 57 -column 8, line 17 -----	1-14, 22-26, 29-53, 59-62, 66-74, 76-78

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International Application No

PCT/GB 99/03161

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0664526 A	26-07-1995	JP 7284129 A	27-10-1995

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**